OPEN ACCESS

University of BRISTOL

(2019). Protein-coding variants contribute to the risk of atopic dermatitis and skin-specific gene expression. *Journal of Allergy and Clinical Immunology*. https://doi.org/10.1016/j.jaci.2019.10.030

Peer reviewed version

Link to published version (if available):
10.1016/j.jaci.2019.10.030

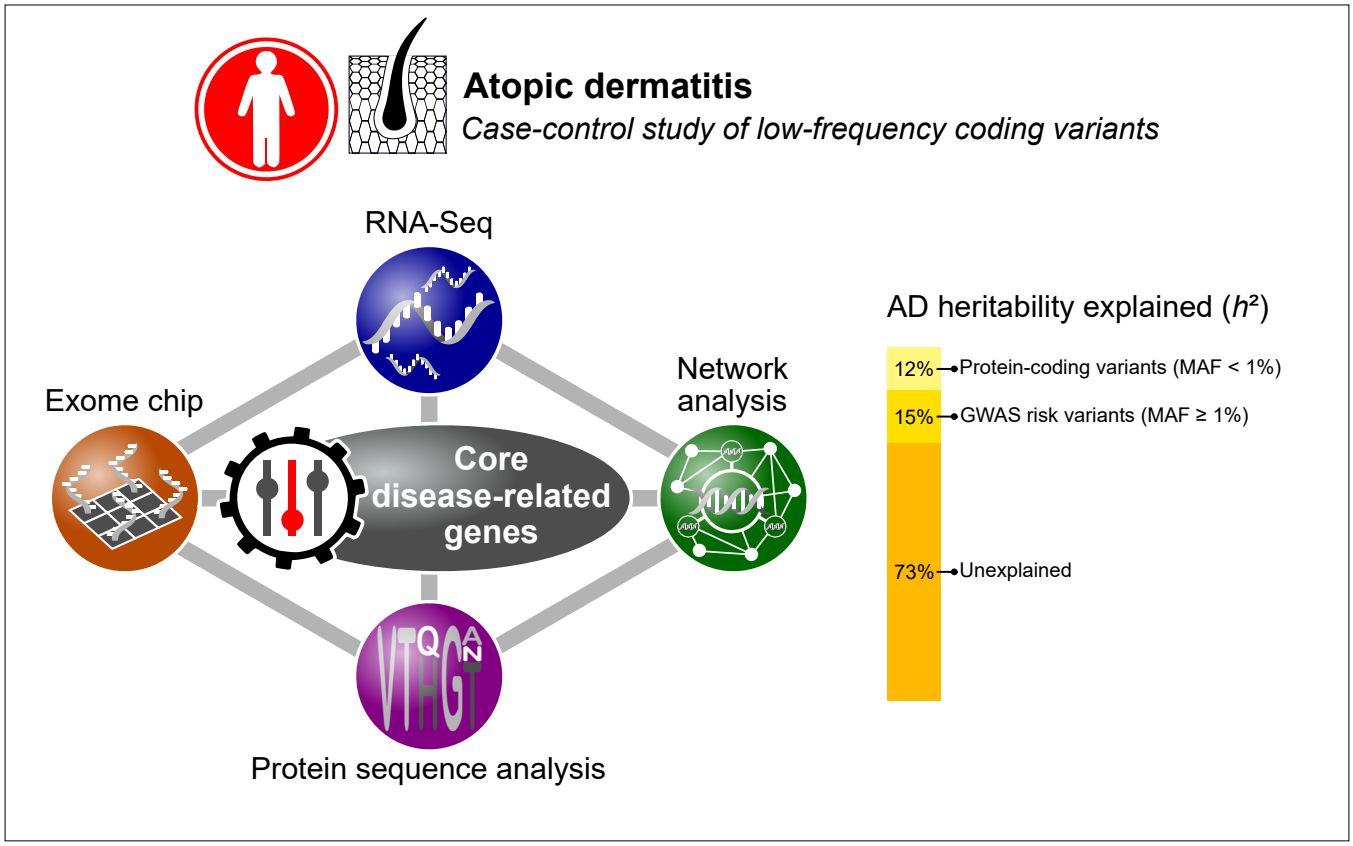
Link to publication record in Explore Bristol Research
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at https://www.sciencedirect.com/science/article/pii/S0091674919314800. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research
General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/

Graphical Abstract



Protein-coding variants contribute to the risk of atopic dermatitis and skin-specific gene expression

Sören Mucha[1,*], MSc, Hansjörg Baurecht[2,3,*], PhD, Natalija Novak[4], MD, Elke Rodríguez[2], PhD, Saptarshi Bej[5], MSc, Gabriele Mayr[1], PhD, Hila Emmert[2], MD, Dora Stölzl[2], MD, Sascha Gerdes[2], MD, Frauke Degenhardt[1], MSc, Matthias Hübenthal[1,2], PhD, Eva Ellinghaus[1], PhD, Eun Suk Jung[1,6], MD, Jan Christian Kässens[1], PhD, Lars Wienbrandt[1], PhD, Wolfgang Lieb[7], MD, Martina Müller-Nurasyid[8,9,10], PhD, Melanie Hotze[2], PhD, Nick Dand[11], PhD, Sarah Grosche[12,13], PhD, Ingo Marenholz[12,13], MD, Andreas Arnold[14], MD, Georg Homuth[15], PhD, Carsten O. Schmidt[16], MD, Ulrike Wehkamp[2], MD, Markus M. Nöthen[17], MD, Per Hoffmann[17], PhD, Lavinia Paternoster[18], PhD, Marie Standl[19], PhD, on behalf of the Early Genetics and Lifecourse Epidemiology (EAGLE) Eczema Consortium, Klaus Bønnelykke[20], MD, Tarunveer S. Ahluwalia[20,21], PhD, Hans Bisgaard[20], Annette Peters[19], PhD, Christian Gieger[22], PhD, Melanie Waldenberger[22], PhD, Holger Schulz[19], PhD, Konstantin Strauch[8,9], PhD, Thomas Werfel[21], MD, Young-Ae Lee[10,11], PhD, Markus Wolfien[5], MSc, Philip Rosenstiel[1], MD, Olaf Wolkenhauer[5], PhD, Stefan Schreiber[1,22], MD, Andre Franke[1,†], PhD, Stephan Weidinger[2,†], MD, David Ellinghaus[1,†], PhD

[1]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany.

[2]Department of Dermatology, Venereology and Allergy, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany.

[3]Department for Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany

[4]Department of Dermatology and Allergy, University Hospital Bonn, Bonn, Germany.

[5]Department of Systems Biology and Bioinformatics, University of Rostock, Germany.

[6]Department of Internal Medicine and Institute of Gastroenterology, Yonsei University College of Medicine, Seoul, Republic of Korea.

[7]Institute of Epidemiology and Biobank PopGen, Christian-Albrechts-University of Kiel, Kiel, Germany.

[8]Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.

[9]Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany.

[10]Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany.

[11]School of Basic & Medical Biosciences, Faculty of Life Sciences & Medicine, King's College London, London, UK.

[12]Pediatric Allergology, Experimental and Clinical Research Center, Charité Universitätsmedizin Berlin, Berlin, Germany.

[13]Max-Delbrück-Centrum (MDC) for Molecular Medicine, Berlin, Germany.

[14]Clinic and Polyclinic of Dermatology, University Medicine Greifswald, Greifswald, Germany.

[15]Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine and Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany.

[16]Institute for Community Medicine, Study of Health in Pomerania/KEF, University Medicine Greifswald, Greifswald, Germany.

[17]Institute of Human Genetics, University of Bonn, Bonn, Germany.

[18]Medical Research Council (MRC) Integrative Epidemiology Unit, Bristol Medical School, University of Bristol, Bristol, UK.

School of Social and Community Medicine, University of Bristol, Bristol, UK.

[19]Institute of Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany.

[20]Copenhagen Prospective Studies on Asthma in Childhood (COPSAC), Herlev and Gentofte Hospital, Denmark.

[21]Steno Diabetes Center Copenhagen, Gentofte 2820, Denmark.

[22]Research Unit of Molecular Epidemiology and Institute of Epidemiology II, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany.

[21]Departement of Dermatology and Allergy, Division of Immunodermatology and Allergy Research, Hannover Medical School, Hannover, Germany.

[22]First Medical Department, University Hospital Schleswig-Holstein, Kiel, Germany.

Corresponding authors:

Stephan Weidinger (sweider@dermatology.uni-kiel.de), Tel. +049 431 500 – 21110, Fax: +049 431 500 – 21118, Arnold-Heller-Straße 3, Haus 19, 24105 Kiel, Germany

and

David Ellinghaus (d.ellinghaus@ikmb.uni-kiel.de), Tel. +049 431 500 – 15131, Fax: +049 431 500 – 15168, Rosalind-Franklin-Str. 12, 24105 Kiel, Germany

*These authors contributed equally to this work.

†These authors jointly supervised this work.

Acknowledgements

We thank Tim Steiert for the help with the graphical summary. This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (SysInflame grant 01ZX1606A; GB-XMAP grant 01ZX1709) as well as the German ELIXIR node de.NBI (de.STAIR grant 031L0106C). The project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany`s Excellence Strategy – EXC 2167-390884018. The project received infrastructure support from the DFG Excellence Cluster No. 306 "Inflammation at Interfaces" and the PopGen Biobank (Kiel, Germany). The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by BMBF and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMU innovativ. SHIP (Study of Health in Pomerania) is part of the Community Medicine Research net (CMR) of the University of Greifswald, Germany, which is funded by BMBF (grants 01ZZ9603, 01ZZ0103 and 01ZZ0403) and the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by BMBF (grant 03IS2061A). Exome chip data have been supported by BMBF (grant no. 03Z1CN22) and the Federal State of Mecklenburg-West Pomerania. LP was supported by a UK Medical Research Council fellowship (MR/J012165/1).

Declaration of Interests

No conflict of interest.

Total word count

3,722

Background: 15% of atopic dermatitis liability-scale heritability could be attributed to 31 susceptibility loci identified by genome-wide association studies, with only three of them (*IL13*, *IL6R*, and *FLG*) resolved to protein-coding variants.

Objective: We examined whether a significant portion of unexplained atopic dermatitis heritability is further explained by low-frequency and rare variants in gene coding sequence.

Methods: We evaluated common, low-frequency and rare protein-coding variants using exome chip and replication genotype data of 15,574 patients and 377,839 controls, combined with whole transcriptome data on lesional, non-lesional and healthy skin samples of 27 patients and 38 controls.

Results: Additional 12.56% (s.e. 0.74%) of atopic dermatitis heritability is explained by rare protein-coding variation. We identified *Docking protein 2* (*DOK2*) and *CD200 Receptor 1* (*CD200R1*) as novel genome-wide significant susceptibility genes. Rare coding variants associated with atopic dermatitis are further enriched in five genes (*IL4R*, *IL13*, *JAK1*, *JAK2*, *TYK2*) of the *IL13* pathway, all of which are targets for novel systemic atopic dermatitis therapeutics. Multiomics-based network and RNA-Sequencing analysis revealed *DOK2* as a central hub interacting, among others, with *CD200R1*, *IL6R* and *STAT3*. Multi-tissue gene expression profile analysis for 53 tissue types from GTEx showed that disease-associated protein-coding variants exert their greatest effect in skin tissues.

Conclusion: Our discoveries highlight a major role of rare coding variants in atopic dermatitis acting independently of common variants. Further extensive functional studies are required to detect all potential causal variants and to specify the contribution of novel susceptibility genes *DOK2* and *CD200R1* to overall disease susceptibility.

Key Messages:

• This exome chip analysis for atopic dermatitis identified two novel susceptibility genes (*DOK2, CD200R1*), and signalling through CD200/CD200R1/DOK2 could be an important new regulatory signalling pathway in AD.

• Disease-associated rare coding variants are enriched in five genes of the IL13 pathway, which are targets for novel systemic atopic dermatitis therapeutics in advanced stages of clinical investigations.

• In total 12% of atopic dermatitis heritability is further explained by rare genetic variation thus highlighting the importance to study the impact of rare protein-coding variants for atopic dermatitis.

Capsule summary:

Protein coding variants in *DOK2, CD200R1* and genes of the *IL13* pathway contribute to disease risk in atopic dermatitis and have the potential to target treatment decisions in atopic dermatitis.

Keywords

Atopic dermatitis; exome chip association analysis; network analysis; protein sequence and structural domain analysis; RNA-Seq

Abbreviations

A1 : Minor allele

A2 : Major allele

AA : Amino acid

AD : Atopic dermatitis

AF : Allele frequency

BC : Betweenness centrality

Bp : Basepair

CC : Closeness centrality

Chr : Chromosome

CI : 95 % confidence interval

ClC : Clustering coefficient

D : Node degree

E.S.S : Essential splice site

EAGLE : Early Genetics and Lifecourse Epidemiology

EDTA : Ethylenediaminetetraacetic acid

eQTL : Expression quantitative trait loci

GANI_MED : Greifswald Approach to Individualized Medicine

GCTA : Genome-wide complex trait analysis

GENUFAD : Genetic Studies in Nuclear Families with AD

GEO : Gene Expression Omnibus

GTEx : Genotype-Tissue Expression

GWAS : Genome-wide association study

HNR : Heinz-Nixdorf Recall

HPRD : Human Protein Reference Database

HRC : Haplotype Reference Consortium

KORA : Kooperative Gesundheitsforschung in der Region Augsburg“

LD : Linkage disequilibrium

MAF : Minor allele frequency

MAGMA : Tool for gene analysis and generalized gene-set analysis of GWAS data

MS : Multiple sclerosis

OR : Odds ratio

PBWT : Positional Burrows-Wheeler Transform

PCA : Principal component analysis

Pfam : Protein Families database

PH : Pleckstrin homology domain

PIP : Posterior inclusion probability

pos : Amino acid position

PP : PolyPhen

PRR1 : Prolin-rich region 1

PRR2 : Prolin-rich region 2

PTB : Phosphotyrosine-binding domain

PWMs : Position weight matrices

QC : Quality control

REML : Restricted maximum-likelihood

SHIP : Study of Health in Pomerania

SiPhy : Site-specificPHYlogenetic

SKAT : Sequence kernel association test

SNP : Single nucleotide polymorphism

SNV : Single nucleotide variant

TMM : Trimmed mean of M-values normalization

TPM : Transcripts Per Million

Introduction

Atopic dermatitis (AD; MIM: 603165) is the most common chronic inflammatory skin disorder affecting 15–20 % of children and 5–10 % of adults (~280 million people worldwide), and the leading cause of the non-fatal disease burden conferred by skin conditions(1). Given its high genetic heritability (71-84 % in Europeans)(2), finding causal genes is a crucial step for developing effective preventive and therapeutic approaches for AD. GWAS so far have identified 31 specific genomic regions associated with AD susceptibility(3-9). The reported susceptibility variants are common (n = 31 with MAF ≥ 5 %) and mostly located in non-coding DNA regions of the genome, have rather small effect sizes (odds ratio (OR) < 1.15) and a largely unclear functional significance(9). Notable exceptions are low-frequency null mutations in the gene encoding the epidermal structural protein filaggrin (*FLG*), which lead to a reduction in biologically active filaggrin peptides, and a complex perturbation of skin barrier function(10), as well as common missense variants in the genes encoding the T helper 2 (T_h2) signature cytokine IL13 (*IL13*; rs20541)(11) as well as the IL6 receptor (*IL6R;* rs2228145)(12). Recently, an exome chip based association study of low-frequency variation across all autosomal exons in multiple sclerosis (MS) cohorts of European ancestry led to the detection of low-frequency MS-associated coding variants for four genes that were missed by previous large-scale MS consortium GWAS(13) and that explain another 5% of MS heritability, thus reopening the debate on the contribution of low-frequency and rare variants to disease risk in complex diseases.

To systematically evaluate the contribution of genetic variation to the genetic architecture of AD on the exome-wide scale, particularly protein-altering variants of low or rare-frequency, we profiled 1,913 AD patients and 14,295 controls in two German cohorts using the Illumina HumanExomeBeadchip (exome chip) (*see* ***Supplementary Table E1 in Online Repository*****)**. The exome chip captures approximately 88 % of low-frequency and rare-coding variants (non-synonymous, splice-site and stop altering, MAFs between 0.01 % and 5 %)

present in Europeans(14). Suggestive significant novel associations ($p_{exomechip} < 1\times10^{-5}$) were taken forward to replication genotyping in a third German cohort of 1,789 AD cases and 3,272 controls, a Danish exome chip case-control study of 292 severe AD cases and 650 controls, GWAS association statistics of the EAGLE eczema consortium of 2,298 independent AD cases and 7,802 controls (*see* ***Supplementary Table E1 in Online Repository*****)** as well as association summary statistics of 361,132 individuals from UK Biobank with self-reported information (*see* ***Supplementary Table E1 in Online Repository*****)**. Gene-based tests and Bayesian fine-mapping analysis as well as whole transcriptome RNA-seq, immunohistochemistry and variant protein analyses were conducted to elucidate potential functional consequences of coding variation associated with AD. Finally, we performed multiomics-based network, pathway gene set and gene expression tissue profile analyses, and we quantified the overall contribution of exome chip variation to AD risk.

Methods

The Methods section in this article's Online Repository provides details on the methods used in this study.

Study samples and genotyping

All cases had been diagnosed with AD by a dermatologist except for UK Biobank cases. All participants provided written informed consent and the study was approved by the ethics boards of the participating institutions, in agreement with the Declaration of Helsinki principles.

German discovery cohort 1. German AD patients (n = 1,056) were recruited at the Department of Dermatology at Christian-Albrechts-University Kiel, Department of Dermatology and Allergy at the Technical University of Munich, and the Department of Dermatology and Allergy at the University of Bonn. Data from healthy control individuals (n = 7,026) were obtained from the PopGen biobank(15), the KORA S4 survey (an independent population-based sample from the general population living in the region of Augsburg, southern Germany)(16) and the Heinz-Nixdorf Recall (HNR) cohort(17), Bonn. AD cases as well as controls were genotyped using Illumina HumanExome-12 v1.0 BeadChips (*see* ***Supplementary Table E1 in Online Repository***).

German discovery cohort 2. German AD patients (n = 1,051) were recruited from dermatology clinics in Kiel or Hannover (the University of Kiel and Medizinische Hochschule of Hannover). The AD cases were genotyped using Illumina HumanCoreExome-24 v1.0 A or HumanCoreExome-24 v1.1 A BeadChips. Data from healthy control individuals (n = 8,135) were obtained from the SHIP and SHIP-TREND cohorts (from the Study of Health in Pomerania, a prospective longitudinal population-based cohort study in West Pomerania)(18). All German controls were genotyped using Illumina HumanExome-12 v1.0 BeadChips (*see* ***Supplementary Table E1 in Online Repository***).

G*erman replication cohort*. German AD patients (n = 1,789) were recruited from dermatology clinics in Kiel or Berlin (University Children's Hospital, Charité Universitätsmedizin Berlin, as part of the Genetic Studies in Nuclear Families with AD (GENUFAD) study). Data from healthy control individuals (n = 3,272) were obtained from University Hospital in Kiel and Lübeck at the Institute of Transfusional Medicine (see ***Supplementary Table E1 in Online Repository***).

Danish replication cohort. All Danish AD cases (n = 292) are hospitalized severe cases from the COPSAC eczema REGISTRY. Healthy control individuals (n = 650) were obtained from the COPSAC2000 and COPSAC2010 birth cohorts(19) in Copenhagen, Denmark. Both, cases and controls were genotyped on the Illumina Infinium OmniExpressExome-8 v1.4 BeadChip (see ***Supplementary Table E1 in Online Repository***).

EAGLE GWAS replication cohorts. We used imputed summary statistics of the EAGLE Eczema Consortium for the discovery cohorts (excluding 23andMe). (https://data.bris.ac.uk/data/dataset/28uchsdpmub118uex26ylacqm) comprising 11,294,660 SNP markers with MAF ≥ 1 % and 10,788 AD cases and 30,047 controls (see ***Supplementary Table E1 in Online Repository***).(9) For replication analysis, we used only European studies independent from our German discovery cohorts and where AD diagnosis was ascertained by a dermatologist.

UK Biobank. Since only 33 (primary diagnosis; field 41203) and 50 patients (secondary diagnosis; field 41204) have been diagnosed with "atopic dermatitis" (ICD10 code L30) in UK Biobank, we used questionnaire information from UK Biobank (release of March 2018). Key words "atopic dermatitis" and/or "eczema" are contained in self-reported data-fields" Non-cancer illness code; self-reported: eczema/dermatitis" (data-field 20002) and "Age hay fever, rhinitis or eczema diagnosed" (data-field 3761) and three others (data-fields 6152, 41202, 41203). For the three lead variants (*DOK2*: rs34215892; rs56094005; *CD200R1*: rs9865242), we downloaded imputed summary association statistics (http://www.nealelab.is/uk-biobank/;

http://www.nealelab.is/uk-biobank/faq; release March 2018; see ***Supplementary Table E1 in Online Repository***) for (a) 83,407 cases with eczema, allergic rhinitis or/and hayfever (data-field 20002) versus 277,120 controls and (b) 9,312 cases with eczema or dermatitis (data-field 3761) versus 351,820 controls.

An overview of the study design is shown in ***Supplementary Figure E1 in Online Repository***, and detailed characteristics of the discovery and replication case-control cohorts are provided in ***Supplementary Tables E1 in Online Repository***.

Results

[Exome chip single-variant association analysis]

In our exome chip discovery search for common, low-frequency and rare variant associations, two German AD discovery cohorts were combined via a meta-analysis resulting in single variant association analysis score statistics of 143,884 genotyped and 1,357,289 single nucleotide variants after imputation with info score ≥ 0.5 (*see* ***Supplementary Figure E1-2, Supplementary Table E2 in Online Repository,*** *see* ***Methods in Online Repository***). 438 and 1,331 SNPs within 7 and 25 loci were identified with genome-wide significance ($p_{exomechip} < 5\times10^{-8}$) and suggestive association ($p_{exomechip} < 1\times10^{-5}$), respectively (see also ***Supplementary Figure E3, Supplementary Table E3 in Online Repository***). To identify novel susceptibility variants outside of established AD GWAS loci (see ***Supplementary Table E4 in Online Repository***) eleven suggestively associated SNPs ($p_{exomechip} < 10^{-5}$ with MAF ≥ 1 %) were selected based on LD clumping method (*see* ***Methods in Online Repository***) and carried forward for replication (***Supplementary Table E5 in Online Repository***). Using the genome-wide threshold of 5×10^{-8} for the combined analysis of discovery and replication (*see* ***Methods in Online Repository***), we identified a novel low frequency missense variant in exon 3 of the gene *Docking protein 2* (*DOK2* at 8p21.3; rs34215892 (p.P274L); $p_{exomechip} = 9.83\times10^{-7}$; genotyped variant) which consistently and robustly replicated in three independent cohorts ($p_{German} = 3.75\times10^{-4}$; $p_{Denmark} = 7.60\times10^{-3}$; $p_{EAGLE} = 4.35\times10^{-2}$; $p_{combined} = 2.15\times10^{-10}$; $OR_{combined} = 0.64$; ***Table 1;*** see ***Supplementary Table E5 in Online Repository***). Further, we detected a novel association between AD and a common missense variant at 3q13.2 (rs9865242; p.E312Q; $p_{combined} = 1.17\times10^{-7}$; $OR_{combined} = 1.16$; genotyped variant; *see* ***Supplementary Table E5 in Online Repository***) located in exon 7 of *CD200 Receptor 1* (*CD200R1*) and 266.51 kb upstream of the previously reported intergenic locus 3q13.2 (*CCDC80*, rs12634229) described only in a Japanese AD cohort(6) so far ($r^2_{rs9865242\text{-}rs12634229} = 0.005$). A look-up of association results from UK Biobank for the self-reported broad allergic disease phenotype "AD (eczema),

allergic rhinitis and/or hayfever"(20) (*see* ***Methods in Online Repository***) further confirmed association signals for *DOK2* (rs34215892; $p_{UK\text{-}Biobank} = 3.35\times10^{-6}$) and *CD200R1* (rs9865242; $p_{UK\text{-}Biobank} = 1.35\times10^{-8}$) (***Table 1***).

[Exome chip gene-based association analysis]

Due to insufficient statistical power to perform single marker tests for rare variants, we performed gene-based association analysis, in which we evaluated the cumulative effects of low-frequency and rare variants (non-synonymous, stop-gain and essential splice site; n = 118,816 exome chip variants excluding imputed variants) for each gene from autosomes (*see* ***Methods in Online Repository***). Only *DOK2* met the exome-wide significance threshold (*see* ***Methods in Online Repository***) and exhibited a stronger association signal than compared to single variant analysis ($p_{combined\text{-}DOK2} = 4.23\times10^{-13}$; ***Table 2;*** *see* ***Supplementary Table E6 in Online Repository***) comprising twelve protein-altering variants of which two were of low-frequency and LD-independent (rs34215892 and rs56094005; 1 % ≤ MAF < 5 %) and ten were rare (MAF < 1 %) (***Figure 1***). We genotyped the second low-frequency variant rs56094005 (p.L138S; $p_{rs56094005\text{-}exomechip} = 4.31\times10^{-3}$; genotyped variant) in our replication set, in addition to rs34215892, and successfully confirmed the single SNP ($p_{rs56094005\text{-}replication} = 5.96\times10^{-3}$; *see* ***Supplementary Table E5 in Online Repository*****)** and the aggregated *DOK2* association signal ($p_{DOK2\text{-}replication} = 1.54\times10^{-6}$; *see* ***Supplementary Table E6 in Online Repository***).

[Bayesian fine-mapping and functional annotation]

As a next step towards understanding the functional causality of the identified AD-associated variants, we carried out Bayesian fine-mapping and functional annotation analysis for loci *CD200R1* and *DOK2* using imputed exome chip data of our discovery cohorts (*see* ***Methods in Online Repository***). Fine-mapping strengthened our hypothesis that lead variants rs9865242 (*CD200R1*) and rs34215892 (*DOK2*) as being most likely causal (in the context of

fine-mapping), with posterior probabilities of 97.2 % and 44.3 %, respectively (*see* ***Supplementary Figure E4 in Online Repository***). rs34215892 overlaps enhancer histone marks and DNase peaks in 15 and 12 different tissues respectively, in each case including the skin, and is predicted to affect protein binding and regulatory motifs (*see* ***Supplementary Table E7 in Online Repository).*** The second *DOK2* lead variant rs56094005 (***Figure 1***) overlaps promoter histone marks and DNase peaks in ten tissues, in each case including T and B cells, and is in perfect LD ($r^2 = 1$) with an intronic variant rs118162691 ($p_{rs118162691\text{-}exomechip} = 4.17\times10^{-3}$; MAF = 3.9 %) which overlaps enhancer histone marks in nine different tissues including the skin (*see* ***Supplementary Table E7 in Online Repository)***. The *CD200R1* missense variant rs9865242 has been suggested as a cis-eQTL for *GTPBP8* in whole blood(21). Ten variants in perfect LD with rs9865242, further overlapping enhancer histone marks in blood predicted to alter regulatory motifs, are located in a conserved region or are found as eQTLs for *CD200R1*(22) (*see* ***Supplementary Table E7 in Online Repository***).

[Immunohistochemistry and whole transcriptome mRNA-seq data analysis]

Immunohistochemistry was used to determine the location of DOK2 in skin tissue (*see* ***Methods in Online Repository***). It showed strong epidermal staining with clear differences among AD lesional, AD non-lesional and healthy skin (*see* ***Supplementary Figure E5 in Online Repository***). DOK2 is predominantly expressed in lymphocytes and the increased abundance of DOK2 in lesional AD skin correlates with the degree of lymphocyte infiltration (*see* ***Supplementary Figure E5 in Online Repository***). Moreover, we observed a significantly increased *DOK2* and *CD200R1* mRNA expression in whole transcriptome mRNA-seq data (*see* ***Methods in Online Repository***) on lesional as compared to non-lesional skin samples of 27 AD patients ($p_{DOK2} = 4.2\times10^{-5}$, $p_{CD200R1} = 2.2\times10^{-5}$) and as compared to skin from 38 healthy individuals ($p_{DOK2} = 8.8\times10^{-11}$, $p_{CD200R1} = 2.2\times10^{-7}$), as well as in AD non-lesional skin compared to healthy skin ($p_{DOK2} = 4.5\times10^{-3}$, $p_{CD200R1} = 2.0\times10^{-2}$) (*see* ***Supplementary Figure***

E6a-b in Online Repository). We also observed a trend of increased expression of *DOK1*, whose protein is a heterodimeric partner for DOK2, in lesional skin compared to non-lesional ($p_{DOK1} = 1.1\times10^{-1}$) or healthy skin ($p_{DOK1} = 3.1\times10^{-4}$), as well as non-lesional skin compared to healthy skin ($p_{DOK1} = 4.0\times10^{-3}$) (*see* ***Supplementary Figure E6c-d in Online Repository***).

[*In silico* variant protein analysis]

To construct a first hypothetical model of whether *CD200R1* and *DOK2* missense lead variants are likely to interfere with functionally active domains on the protein level, we performed extensive literature search and further conducted protein domain analyses of DOK2 and the CD200/CD200R1 receptor complex (*see* ***Methods in Online Repository***) (***Figure 2***). Members of the DOK adapter protein family act as regulators of cell stimulatory signals by serving as substrates for diverse receptor and cytoplasmic kinases, and the highly similar and interacting members DOK1 and DOK2 are involved in the down-regulation of immune receptor signalling in $CD4^+$ T cells as well as myeloid cells such as macrophages and neutrophils(23). It is assumed that the inhibitory role of DOK2 is accomplished by recruiting and activating RasGAP to inhibit Ras and thus to suppress pro-inflammatory ERK, JNK and MAPK pathways for the DOK2 response to CD200R1 in human myeloid cells(24). Activated RasGAP inhibits mitogen-activated kinase (MAPK) signalling and subsequently reduces production of pro-inflammatory cytokines such as TNFα, INFγ, IL1, IL17, IL6, IL8(25, 26). By means of our protein sequence and structure analyses we observed that variant rs56094005 (p.L138S) locates within a linker sequence between the PH and the PTB domain adjacent the Y139 phosphorylation-dependent DOK1 interaction site and may interfere with heterodimerization of DOK1 and DOK2 required for full phosphorylation of the two proteins and signalling(27). Variant rs34215892 (p.P274L), located in the invariant RasGAP-SH2 binding consensus motif YxxPxD(27), is likely affecting local protein structure conformation due to the unique structural rigidity of the proline side chain and therefore predicted to interfere with RasGAP signalling.

Assuming an important stabilizing structural role of the proline within the binding motif, the variant p.P274L is likely to disturb the RasGAP activation by DOK2. The amino acid substitution of E to Q of variant rs9865242 (p.E312Q) causes a loss of negative charge in proximity of the NPLY motif and protein interaction site and may therefore modify protein-protein contacts and signalling(28).

[AD core network construction and differential gene expression analysis of core genes]

In order to assess possible interactions of genes *CD200R1* and *DOK2* with candidate genes from AD GWAS loci(3-9), we generated an AD core interaction network using network prioritization algorithms which make use of protein-protein, protein-gene, co-expression, and shared protein domains data from public repositories (*see* ***Supplementary Figure E7; Supplementary Table E8 in Online Repository***; *see* ***Methods in Online Repository***). We identified *DOK2* as a central hub node interacting with *CD200R1* and functionally established AD susceptibility genes *STAT3*(9), *MICB*(29), *CLEC16A*(8) and *IL6R*(12) (***Figure 3a***). We used our aforementioned whole transcriptome mRNA-seq data (*see* ***Methods in Online Repository***) to assess whether expression levels from genes of our AD core network are differentially expressed in lesional, non-lesional and healthy skin from AD patients and healthy individuals. 22 out of the 30 AD core genes (***Figure 3a***) are significantly up or downregulated (*see* ***Supplementary Table E9; Supplementary Figure E8 in Online Repository***), with 17 out of 22 directly interacting with *DOK2* (***Figure 3b***), thus indicating the biological importance of *DOK2* in AD pathogenesis.

[Pathway and gene expression tissue specificity analysis for exome chip variants]

To reveal potential differences in terms of biological pathways and involved tissues for exome chip variant (exome chip) and common variants investigated in AD GWAS meta-analysis studies (GWAS) from the EAGLE consortium, we performed gene-set pathway

enrichment analysis for curated gene sets and gene ontology (GO) terms as well as tissue-specific gene expression profile analysis(30) for 53 tissue types and 11,688 samples from GTEx release 7 (*see* ***Methods in Online Repository***). The pathway enrichment analysis identified one significant gene set including five genes ($P_{Bonferroni} < 0.05$; *IL4R*, *IL13*, *JAK1*, *JAK2*, *TYK2*) of the *IL13* pathway for the exome chip in comparison to eight blood-cell related gene sets (including regulation of immunoglobulin production, B cell activation and B cell mediated immunity) for the EAGLE GWAS data. Recently, AD was characterized as an IL13-dominant disease based on high-depth RNA-seq transcriptome data of 147 samples from cohorts of AD patients, psoriasis patients, and healthy controls, with IL13 being the most distinctive marker for AD(31). We hypothesize that low-frequency and rare coding variants in genes of the *IL13* pathway are further likely to be associated with AD. In the tissue specificity analysis, we tested for relationships between tissue-specific gene expression profiles and variant association statistics from exome chip and GWAS, respectively. For the exome-chip data, we observed a Bonferroni-corrected significant association ($P < 9.43\times10^{-4}$) with two skin tissue types (sun exposed and non-sun exposed skin; ***Figure 4a***). In comparison the EAGLE GWAS data revealed a significant association with tissues whole blood and spleen (***Figure 4b***). Thus, exome chips variants cumulatively may have a stronger effect on skin tissue gene expression than common variants interrogated in AD consortium GWAS analyses.

[Estimation of liability-heritability from exome chip variants]

To quantify the overall contribution of exome chip variation to AD risk, i.e. estimating the exome-wide contribution of (mainly rare) protein-coding variants to the liability-scale heritability for AD, we used a restricted maximum-likelihood approach to model heritability attributable to genotypic variation across genotyped-only exome chip variants (*see* ***Methods in Online Repository***). We found that common variants from exome chip (LD pruned, MAF ≥ 5 %; *see* ***Methods in Online Repository***) explain 1.04 % of heritability on the observed

scale (corresponding to 2.93 % on the liability scale assuming a prevalence of 0.14). By lowering the MAF threshold to 1%, exome chip heritability increased to 1.27 % (3.60 %). Dividing variants into low-frequency (1 % ≤ MAF < 5 %) and rare (MAF < 1 %), heritability was found to be 0.37 % for low-frequency variants (0.94 %) and 4.47 % for rare variants (12.56 %). Furthermore, we estimated the heritability explained by the newly identified lead variants of *DOK2* (rs3215892, rs56094005) and *CD200R1* (rs9865242). Variants rs3215892, rs56094005 and rs9865242 explain approximately 0.015 % (0.041 %), 0.004 % (0.012 %) and 0.013 % (0.035 %) on the observed scale (liability scale), respectively.

Discussion

In conclusion, we analysed the association of AD with common, low-frequency and rare protein-coding variants and implicate two novel genes (*DOK2* and *CD200R1*) contributing to AD risk. So far, 14.91 % (6.95 % excluding *FLG* mutations) of AD liability-scale heritability could be attributed to common lead variants of 31 GWAS loci, estimated from a set of 5,775 clinically diagnosed AD patients and assuming a population prevalence of 0.15(3-9). Recently, Ferreira *et al*.(20) further reported a genome-wide liability-scale heritability estimate of 9.04 % from 1.2 million HapMap SNPs (assuming a population prevalence of 0.14) and on the basis of self-reported information from questionnaires. Our newly identified variants (rs3215892, rs56094005, rs9865242) explain another 0.088 % of the variance in liability. Interestingly, approximately 12.56 % (s.e. 0.74 %) of liability-scale heritability is estimated to be explained by rare protein-coding variants (MAF < 1 %), thus highlighting the importance to study rare coding variation in AD. Since these heritability estimates originate from predominantly European study populations, performing GWAS and exome chip studies in more diverse populations is crucial to optimize heritability estimation for AD in the future. Our results encourage future studies along the same path and highlight the importance to study the impact of protein-coding variants for phenotypically well-defined clinically diagnosed cohorts.

Pathway enrichment analysis of exome chip data revealed that association signals of low-frequency and rare coding variants are enriched in five genes of the *IL13* pathway, all of which are targets for novel systemic AD therapeutics in advanced stages of clinical investigations(32), supporting the pivotal role of type 2 inflammation in AD pathogenesis(1). Gene expression tissue profile analysis showed that exome chip variants associated with AD cumulatively have a stronger effect on skin tissue gene expression than common GWAS variants associated with AD, as identified in combination with whole transcriptome RNA-sequencing data on lesional, non-lesional and healthy skin tissue as well as tissue samples from GTEx. In accordance with results from the exome chip study for multiple sclerosis(13), we

observed that the minor allele of low-frequency lead missense variants in *DOK2* is mostly protective (in context of the odds ratio; ***Table 2***).

The newly discovered genes *DOK2* and *CD200R1* have clear immunological functions, confirming that AD pathogenesis is primarily driven by immune dysregulation. Structural protein domain analysis, topological network and differential gene expression analyses suggest that missense variants in *DOK2* (rs34215892; rs56094005) in combination with the missense AD risk variant in *CD200R1* (rs9865242) together may affect tyrosine phosphorylation sites in DOK2 and CD200R1 (***Figure 2***). *DOK2* belongs to the DOK gene family encoding for seven different DOK proteins (DOK1–7) being involved in signal transduction(27, 33-35). DOK1–2 are adaptor proteins and mainly expressed in hematopoietic/immune cells(36) and have been implicated to negatively regulate proliferation and constitutive expression of DOK2(37). Further studies showed that both DOK1–2 are essential negative regulators of ERK signalling downstream of Toll-like receptor 4(38). Mice lacking DOK1–3 showed significant defects in immune cell development and in immune responses(39). Furthermore, DOK1–2 play a role in the maturation of NK cells(40), which have been shown to be reduced in AD. In line with these observations, we observed a significantly increased *DOK2* and *CD200R1* mRNA expression in lesional as compared to non-lesional skin samples of AD patients, and as compared to skin from healthy individuals. Our AD core network (***Figure 3***) further showed that *DOK2* acts as a central hub gene interacting with *CD200R1* as well as several candidate AD GWAS susceptibility genes on the cellular level. Our AD core network revealed 21 genes that are directly functionally related to *DOK2* of which 16 are significantly upregulated (including *DOK1*) and one (*RASGAP*) is significantly downregulated in lesional skin samples of AD patients. Taking into account the reported inflammatory role of both DOK2 and CD200R1, we conclude that signalling through CD200/CD200R1/DOK2 could be an important new regulatory signalling pathway in AD. Extensive functional studies are required to detect all

potential causal variants and thus to specify the contribution of the *DOK2* and *CD200R1* to overall disease susceptibility.

Author contributions

A.F., S.W. and D.E conceived and designed the study. A.F. and D.E. supervised the study. W.L., M.M-N., M.M.N., P.H., A.A., G.H., C.O.M., M.S., K.B., T.S.A., H.Bi., D.S., S.G., A.P., C.G., M.Wa., H.S., T.W., E.R., K.S., T.W., R.F-H., N.N., M.Ho., P.R., S.S., Y.-A. L., and S.W. contributed to sample collection and phenotyping. S.M., D.E., E.R. and E.E. executed sample handling, genotyping and/or quality control. U.W. performed evaluation of immunohistochemistry. G.M. performed *in silico* variant protein analysis. S.M. and H.B. performed statistical and bioinformatics analysis for the discovery and the combined meta analyses, while T.S.A. performed statistical analysis for the Danish replication cohort and S.G., I.M. performed statistical analysis for UK Biobank. H.E., F.D., M.Hü., J.C.K., L.W. and D.E. helped with bioinformatic analyses. S.B., M.W., O.W., S.M. and D.E. performed network analysis. S.M., H.B., G.M., E.E., E.S.J., N.D., L.P., S.W., A.F. and D.E. participated in interpretation of results. S.M., H.B. and D.E. wrote the manuscript. All authors critically reviewed and approved the manuscript.

References

1. Weidinger S, Novak N. Atopic dermatitis. Lancet. 2016;387(10023):1109-22.
2. Ober C, Yao TC. The genetics of asthma and allergic disease: a 21st century perspective. Immunol Rev. 2011;242(1):10-30.
3. Schaarschmidt H, Ellinghaus D, Rodriguez E, Kretschmer A, Baurecht H, Lipinski S, et al. A genome-wide association study reveals 2 new susceptibility loci for atopic dermatitis. J Allergy Clin Immunol. 2015;136(3):802-6.
4. Esparza-Gordillo J, Weidinger S, Folster-Holst R, Bauerfeind A, Ruschendorf F, Patone G, et al. A common variant on chromosome 11q13 is associated with atopic dermatitis. Nat Genet. 2009;41(5):596-601.
5. Sun LD, Xiao FL, Li Y, Zhou WM, Tang HY, Tang XF, et al. Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. Nat Genet. 2011;43(7):690-4.
6. Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Sakashita M, et al. Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. Nat Genet. 2012;44(11):1222-6.
7. Paternoster L, Standl M, Chen CM, Ramasamy A, Bonnelykke K, Duijts L, et al. Meta-analysis of genome-wide association studies identifies three new risk loci for atopic dermatitis. Nat Genet. 2011;44(2):187-92.
8. Ellinghaus D, Baurecht H, Esparza-Gordillo J, Rodriguez E, Matanovic A, Marenholz I, et al. High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. Nat Genet. 2013;45(7):808-12.
9. Paternoster L, Standl M, Waage J, Baurecht H, Hotze M, Strachan DP, et al. Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat Genet. 2015;47(12):1449-56.
10. Irvine AD, McLean WH, Leung DY. Filaggrin mutations associated with skin and allergic diseases. N Engl J Med. 2011;365(14):1315-27.
11. Tsunemi Y, Saeki H, Nakamura K, Sekiya T, Hirai K, Kakinuma T, et al. Interleukin-13 gene polymorphism G4257A is associated with atopic dermatitis in Japanese patients. J Dermatol Sci. 2002;30(2):100-7.
12. Esparza-Gordillo J, Schaarschmidt H, Liang L, Cookson W, Bauerfeind A, Lee-Kirsch MA, et al. A functional IL-6 receptor (IL6R) variant is a risk factor for persistent atopic dermatitis. J Allergy Clin Immunol. 2013;132(2):371-7.
13. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al. Low-Frequency and Rare-Coding Variation Contributes to Multiple Sclerosis Risk. Cell. 2018;175(6):1679-87 e7.
14. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.
15. Nothlings U, Krawczak M. [PopGen. A population-based biobank with prospective follow-up of a control group]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2012;55(6-7):831-5.
16. Wichmann HE, Gieger C, Illig T, Group MKS. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. Gesundheitswesen. 2005;67 Suppl 1:S26-30.
17. Schmermund A, Mohlenkamp S, Stang A, Gronemeyer D, Seibel R, Hirche H, et al. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. Am Heart J. 2002;144(2):212-8.

18. Volzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N, et al. Cohort profile: the study of health in Pomerania. Int J Epidemiol. 2011;40(2):294-307.
19. Bisgaard H. The Copenhagen Prospective Study on Asthma in Childhood (COPSAC): design, rationale, and baseline data from a longitudinal birth cohort study. Ann Allergy Asthma Immunol. 2004;93(4):381-9.
20. Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet. 2017;49(12):1752-7.
21. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45(10):1238-43.
22. Fehrmann RS, Jansen RC, Veldink JH, Westra HJ, Arends D, Bonder MJ, et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS Genet. 2011;7(8):e1002197.
23. Mashima R, Hishida Y, Tezuka T, Yamanashi Y. The roles of Dok family adapters in immunoreceptor signaling. Immunol Rev. 2009;232(1):273-85.
24. Mihrshahi R, Barclay AN, Brown MH. Essential roles for Dok2 and RasGAP in CD200 receptor-mediated regulation of human myeloid cells. J Immunol. 2009;183(8):4879-86.
25. Holmannova D, Kolackova M, Kondelkova K, Kunes P, Krejsek J, Andrys C. CD200/CD200R paired potent inhibitory molecules regulating immune and inflammatory responses; Part I: CD200/CD200R structure, activation, and function. Acta Medica (Hradec Kralove). 2012;55(1):12-7.
26. Zhang S, Cherwinski H, Sedgwick JD, Phillips JH. Molecular mechanisms of CD200 inhibition of mast cell activation. J Immunol. 2004;173(11):6786-93.
27. Boulay I, Nemorin JG, Duplay P. Phosphotyrosine binding-mediated oligomerization of downstream of tyrosine kinase (Dok)-1 and Dok-2 is involved in CD2-induced Dok phosphorylation. J Immunol. 2005;175(7):4483-9.
28. Mihrshahi R, Brown MH. Downstream of tyrosine kinase 1 and 2 play opposing roles in CD200 receptor signaling. J Immunol. 2010;185(12):7216-22.
29. Schrambach S, Ardizzone M, Leymarie V, Sibilia J, Bahram S. In vivo expression pattern of MICA and MICB and its relevance to auto-immunity and cancer. PLoS One. 2007;2(6):e518.
30. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219.
31. Tsoi LC, Rodriguez E, Degenhardt F, Baurecht H, Wehkamp U, Volks N, et al. Atopic dermatitis is an IL-13 dominant disease with greater molecular heterogeneity compared to psoriasis. J Invest Dermatol. 2019.
32. Weidinger S, Beck LA, Bieber T, Kabashima K, Irvine AD. Atopic dermatitis. Nat Rev Dis Primers. 2018;4(1):1.
33. Abramson J, Rozenblum G, Pecht I. Dok protein family members are involved in signaling mediated by the type 1 Fcepsilon receptor. Eur J Immunol. 2003;33(1):85-91.
34. Gerard A, Favre C, Garcon F, Nemorin JG, Duplay P, Pastor S, et al. Functional interaction of RasGAP-binding proteins Dok-1 and Dok-2 with the Tec protein tyrosine kinase. Oncogene. 2004;23(8):1594-8.
35. Yasuda T, Bundo K, Hino A, Honda K, Inoue A, Shirakata M, et al. Dok-1 and Dok-2 are negative regulators of T cell receptor signaling. Int Immunol. 2007;19(4):487-95.
36. Di Cristofano A, Carpino N, Dunant N, Friedland G, Kobayashi R, Strife A, et al. Molecular cloning and characterization of p56dok-2 defines a new family of RasGAP-binding proteins. J Biol Chem. 1998;273(9):4827-30.

37. Gugasyan R, Quilici C, I ST, Grail D, Verhagen AM, Roberts A, et al. Dok-related protein negatively regulates T cell development via its RasGTPase-activating protein and Nck docking sites. J Cell Biol. 2002;158(1):115-25.
38. Shinohara H, Inoue A, Toyama-Sorimachi N, Nagai Y, Yasuda T, Suzuki H, et al. Dok-1 and Dok-2 are negative regulators of lipopolysaccharide-induced signaling. J Exp Med. 2005;201(3):333-9.
39. Mashima R, Arimura S, Kajikawa S, Oda H, Nakae S, Yamanashi Y. Dok adaptors play anti-inflammatory roles in pulmonary homeostasis. Genes Cells. 2013;18(1):56-65.
40. Celis-Gutierrez J, Boyron M, Walzer T, Pandolfi PP, Jonjic S, Olive D, et al. Dok1 and Dok2 proteins regulate natural killer cell development and function. EMBO J. 2014;33(17):1928-40.
41. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28(10):1045-8.
42. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8(1):1826.

Table 1 – Exome chip discovery and replication genotyping as well as UK Biobank replication (self-reported AD) single-marker association analysis revealed *DOK2* and *CD200R1* as genome-wide significant AD susceptibility genes. Since exome chip discovery and replication genotyping studies comprised only AD patients ascertained by a dermatologist in comparison to the UK Biobank replication study with only self-reported AD, association results were listed separately.

dbSNP ID AA substitution	**Position (chr:bp)**	**Gene**	**Variant type**	**A1/ A2**	**AF_{cases}**	**$AF_{controls}$**	**Discovery $n_{cases}/n_{controls}$**	**Discovery p-value/OR (95 % CI)**	**Replication $n_{cases}/n_{controls}$**	**Replication p-value/OR (95 % CI)**	**Combined p-value/OR (95 % CI)**
rs34215892 P274L	8:21767240	*DOK2*	missense	A/G	0.025	0.039	1,913/14,295	9.83×10^{-7}; 0.61 (0.34–0.86)	3,962/10,076	1.35×10^{-5}; 0.68 (0.56–0.79)	**2.15×10^{-10}; 0.64 (0.53–0.76)**
rs9865242 E312Q	3:112642568	*CD200R1*	missense	C/G	0.457	0.435	1,913/14,295	5.80×10^{-6}; 1.18 (1.15–1.22)	4,349/11,724	3.38×10^{-3}; 1.14 (1.09–1.19)	1.17×10^{-7}; 1.16 (1.13–1.19)
					UK Biobank summary statistics for AD, allergic rhinitis and/or hayfever (self-reported)				**UK Biobank summary statistics for AD or dermatitis (self-reported)**		
dbSNP ID AA substitution	**Position (chr:bp)**	**Gene**	**Variant type**	**A1/ A2**	**AF_{cases}**	**$AF_{controls}$**	**Phenotype 1 $n_{cases}/n_{controls}$**	**p-value**	**AF_{cases}**	**$AF_{controls}$**	**Phenotype 2 $n_{cases}/n_{controls}$**	**p-value**
rs34215892 P274L	8:21767240	*DOK2*	missense	A/G	0.027	0.029	83,407/277,120	3.35×10^{-6}	0.025	0.029	9,312/351,820	2.50×10^{-3}
rs9865242 E312Q	3:112642568	*CD200R1*	missense	C/G	0.474	0.467	83,407/277,120	**1.35×10^{-8}**	0.487	0.468	9,312/351,820	8.33×10^{-8}

AA, amino acid; chr, chromosome of the marker; bp, genomic position from NCBI dbSNP build v150 (genome build hg19); Gene, candidate gene; A1, minor allele; A2, major allele; AF, allele frequency of A1; OR, estimated odds ratio; P-values and ORs were calculated with respect to the minor

allele; genome-wide significant p-values ($p < 5\times10^{-8}$) are indicated in bold. All AD cases from discovery and replication panels had been diagnosed with AD (eczema) by a dermatologist. rs34215892 and rs9865242 were genotyped (non-imputed) in discovery and replication panels (except for the EAGLE GWAS replication data). Association results of UK Biobank (self-reported phenotypes; *see* ***Supplementary Table E1 in Online Repository***) are listed separately.

Table 2 – Meta-analysis of gene-based aggregation tests for *DOK2* increased the genome-wide significant association signal by more than two orders of magnitude in comparison to single-marker analysis (*Table 1*), indicating that multiple rare variants (with 10 out of 12 variants predicted to be pathogenic; *Figure 1*) contribute to the association signal.

Gene	dbSNP ID	Position (chr:bp)	AF_{mean}	AA	Prediction	Discovery $n_{cases}/n_{controls}$ $pSKAT_{discovery}$	Replication $n_{cases}/n_{controls}$ $pSKAT_{replication}$	Combined $n_{cases}/n_{controls}$ $pSKAT_{combined}$
DOK2	rs2242241[#]	8:21766881	0.00028	S394A	Tolerated, possibly damaging	1,913/14,295 **2.61×10^{-7}**	3,932/10,076 **1.54×10^{-6}**	5,845/24,371 **4.23×10^{-13}**
	rs145725971[*]	8:21767033	0.00065	H343R	Tolerated, possibly damaging			
	rs145405180[#]	8:21767148	0.00139	A305T	Tolerated, benign			
	rs34215892[#]	8:21767240	0.03708	P274L	Damaging, probably damaging			
	rs74909419[*]	8:21767265	0.00019	R266W	Damaging, probably damaging			
	rs141482665[*]	8:21767322	0.00012	N247D	Tolerated, probably damaging			
	rs200168233[*]	8:21767414	0.00015	R216H	Damaging, probably damaging			
	rs200503110[#]	8:21767417	0.00019	R215H	Damaging, probably damaging			
	rs149080191[*]	8:21768370	0.00009	E.S.S.				
	rs56094005[#]	8:21769432	0.04026	L138S	Damaging, probably damaging			
	rs201025320[#]	8:21770011	0.00003	R25H	Damaging, probably damaging			
	rs142660088[#]	8:21771080	0.00043	L11F	Damaging, probably damaging			

AA, amino acid substitution; E.S.S., essential splice site; Gene, candidate gene; Chr, chromosome of the marker; Bp, genomic position from NCBI dbSNP build v150 (genome build hg19); AF_{mean}, mean allele frequency of minor allele from discovery panels (*see* ***Supplementary Table E5 in Online Repository***); Prediction, SIFT prediction, PolyPhen-2 prediction; pSKAT, p of Sequence Kernel Association Test; Bonferroni-corrected exome chip significant gene-based p-values ($p_{gene} < 0.05/15{,}998 = 3\times10^{-6}$; 15,998 genes) are indicated in bold. Cohort specific association details are given in ***Supplementary Table E6 in Online Repository.*** *, risk variant (in context of the odds ratio); #, protective variant. All 12 variants were genotyped (non-imputed) in the discovery and replication panels (except for the EAGLE GWAS replication data).

Figure 1 – Exome chip association analysis identified two low-frequency (rs34215892 and rs56094005; 1 % ≤ MAF < 5 %) and ten rare (MAF < 1 %) coding variants contributing to the genome-wide significant association signal at *DOK2* (p_{DOK2} = 4.23×10^{-13}; *Table 2*). Minor allele frequency in AD cases (dark blue) and controls (light blue) are shown from discovery exome chip meta-analysis (*see* ***Supplementary Table E1 in Online Repository***). 10 out of 12 variants were predicted to be pathogenic (*see* ***Methods in Online Repository***), suggesting that multiple missense variants contribute to the gene-based association signal. Both low-frequency lead variants are LD-independent ($r^2_{rs56094005\text{-}rs34215892}$ = 0.0015). Variant effect predictions (by SIFT and PolyPhen-2) depicted in red (white or grey) represent amino acid substitutions predicted to be potentially damaging (probably not damaging or prediction not possible). SNV – single nucleotide variant; ESS – essential splice site; pos – amino acid position; PP – PolyPhen-2. PH – Pleckstrin-homology domain; PTB – Phosphotyrosine-binding domain; PRR1 – Prolin-rich region 1; PRR2 – Prolin-rich region 2. A – Alanine; D – Aspatic acid; F – Phenylalanine; H – Histidine; L – Leucin; N – Asparagine; P – Proline; R – Arginine; S – Serine; T – Threonine; W – Tryptophan.

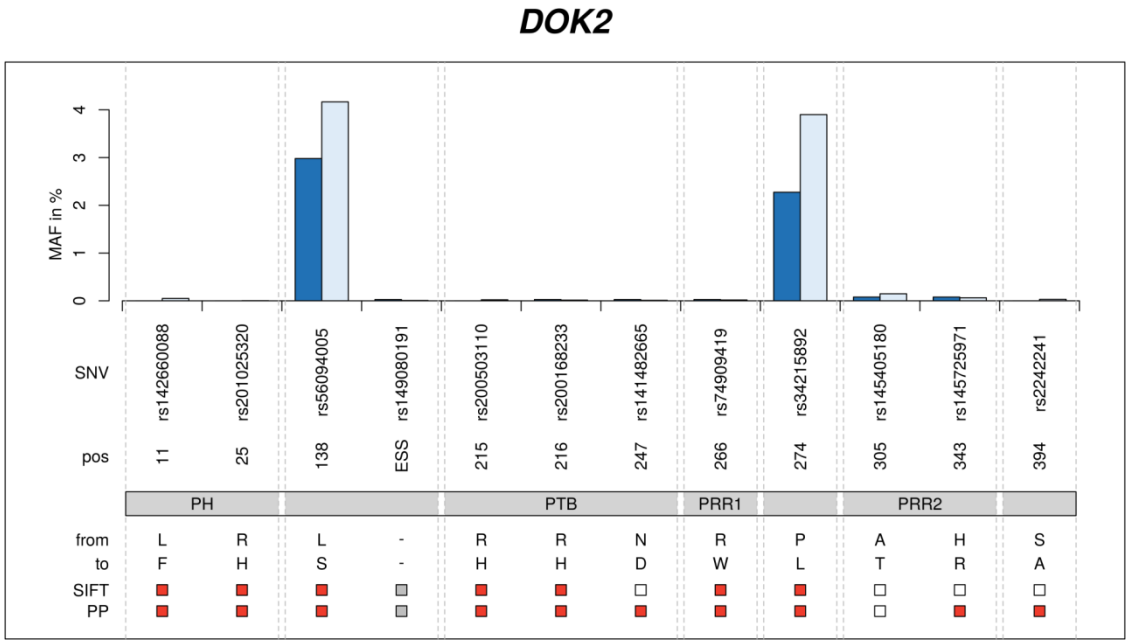


Figure 2 – Hypothetical model constructed from protein sequence and structural domain analysis suggesting that missense lead variants rs56094005, rs34215892 (p.L138S and p.P274L; *DOK2*) and rs9865242 (p.E312Q; *CD200R1*) are located near functionally important tyrosine phosphorylation sites and may interfere with CD200/CD200R1 receptor complex and DOK2 function.

We assume a simplified illustration of DOK2 function in response to CD200R1 in human myeloid cells(41) in which CD200/CD200R1 binding leads to tyrosine phosphorylation of the NPLY motif in the cytoplasmic tail of CD200R1 and the recruitment of the DOK2 adapter protein(24). DOK2 interaction with CD200R1 leads to DOK2 tyrosine phosphorylation at positions Y271/Y299 (activating RasGAP) and Y139 (activating DOK1), which leads to the recruitment of RasGAP and its subsequent activation(28).

Interacting CD200 and CD200R1 extracellular domains are visualized as structural models received from the Protein Databank (PDB ID 4bfi), while CD200R1 transmembrane and cytoplasmic tail is shown as blue lines. Variants are highlighted in red to indicate their relative positions within protein domains. The structural effect of protein variants cannot be modelled in 3D due to the lack of structural templates in the PDB for the concerned regions. Interactions between proteins are visualized as solid lines, following events as dashed lines. Cell membranes anchoring the receptors are illustrated in orange-grey. DOK2 domains were abbreviated as PH: Pleckstrin-homology domain, PTB: Phosphotyrosine-binding domain, PRR: Proline-rich region.

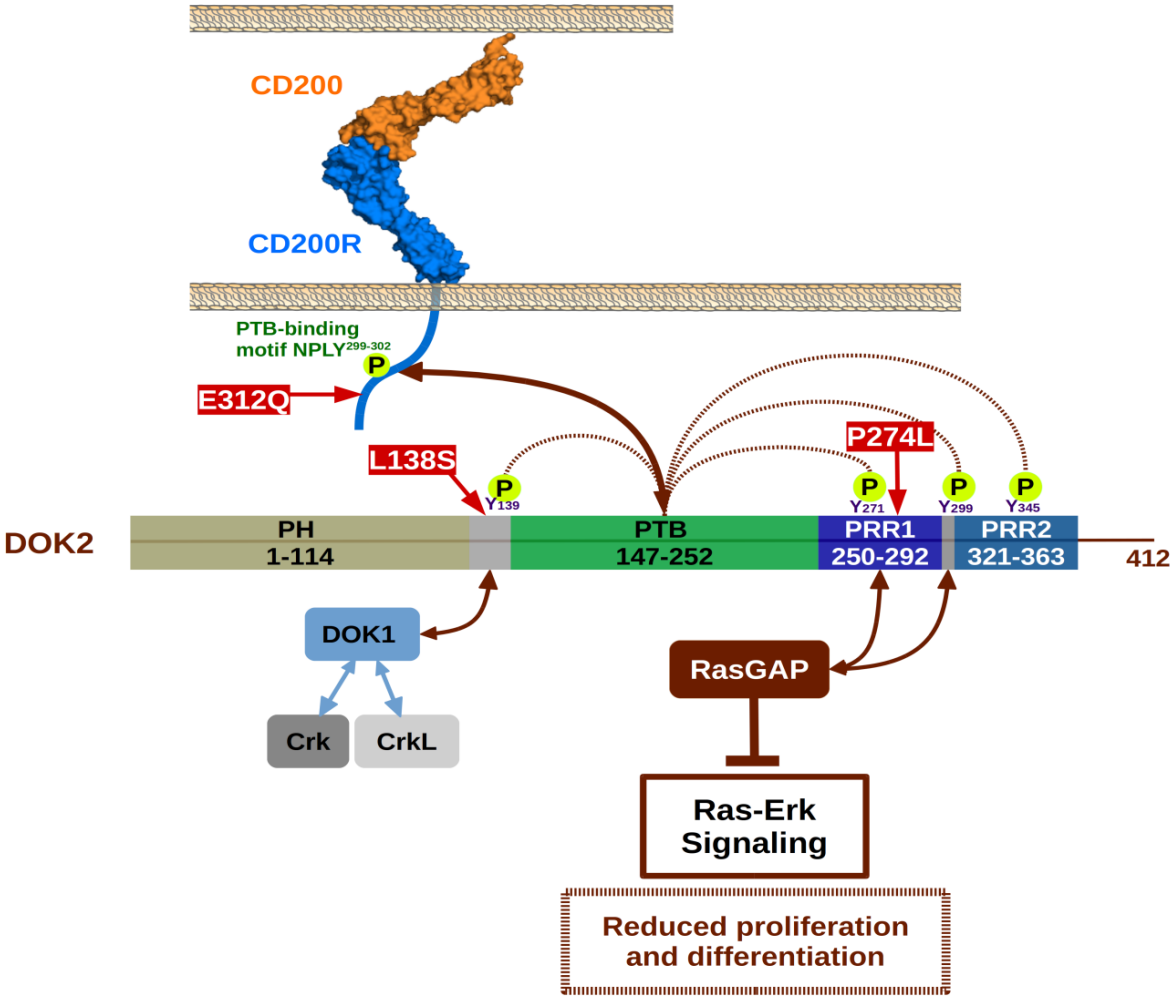

CD200
CD200R
PTB-binding
motif NPLY299-302
P
E312Q
L138S
P274L
P
Y139
P
Y271
P
Y299
P
Y345
DOK2
PH
1-114
PTB
147-252
PRR1
250-292
PRR2
321-363
412
DOK1
Crk
CrkL
RasGAP
Ras-Erk
Signaling
Reduced proliferation
and differentiation

Figure 3 – Multiomics-based network interaction analysis unveiled topologically and functionally important AD susceptibility genes and determined *DOK2* as a central hub node interacting with *CD200R1* and candidate susceptibility genes (marked with white asterisk) identified by previous large AD consortium GWAS. (A) The majority of core genes (n = 22 out of 30; including *DOK2*) shows significantly upregulated and downregulated gene expression levels (depicted by black upwards and downwards arrows) in whole transcriptome mRNA-seq data ("AD-L"; "AD-NL" and "healthy"; see also ***Supplementary Table E9 in Online Repository***). Grey-striped nodes (n = 10, including *DOK2*) represent the most relevant genes from topological network analysis, with two light grey-striped nodes representing genes not interacting with *DOK2* and seven dark grey-striped nodes representing genes directly interacting with *DOK2*. Non-striped nodes (n = 20) represent genes additionally added by a functional similarity search, with six light grey non-striped nodes not directly interacting with *DOK2* and 14 dark grey non-striped nodes directly interacting with *DOK2*. **(B)** 17 out of the significantly up- or downregulated genes, depicted with log2-transformed gene expression counts, are interacting with *DOK2*.

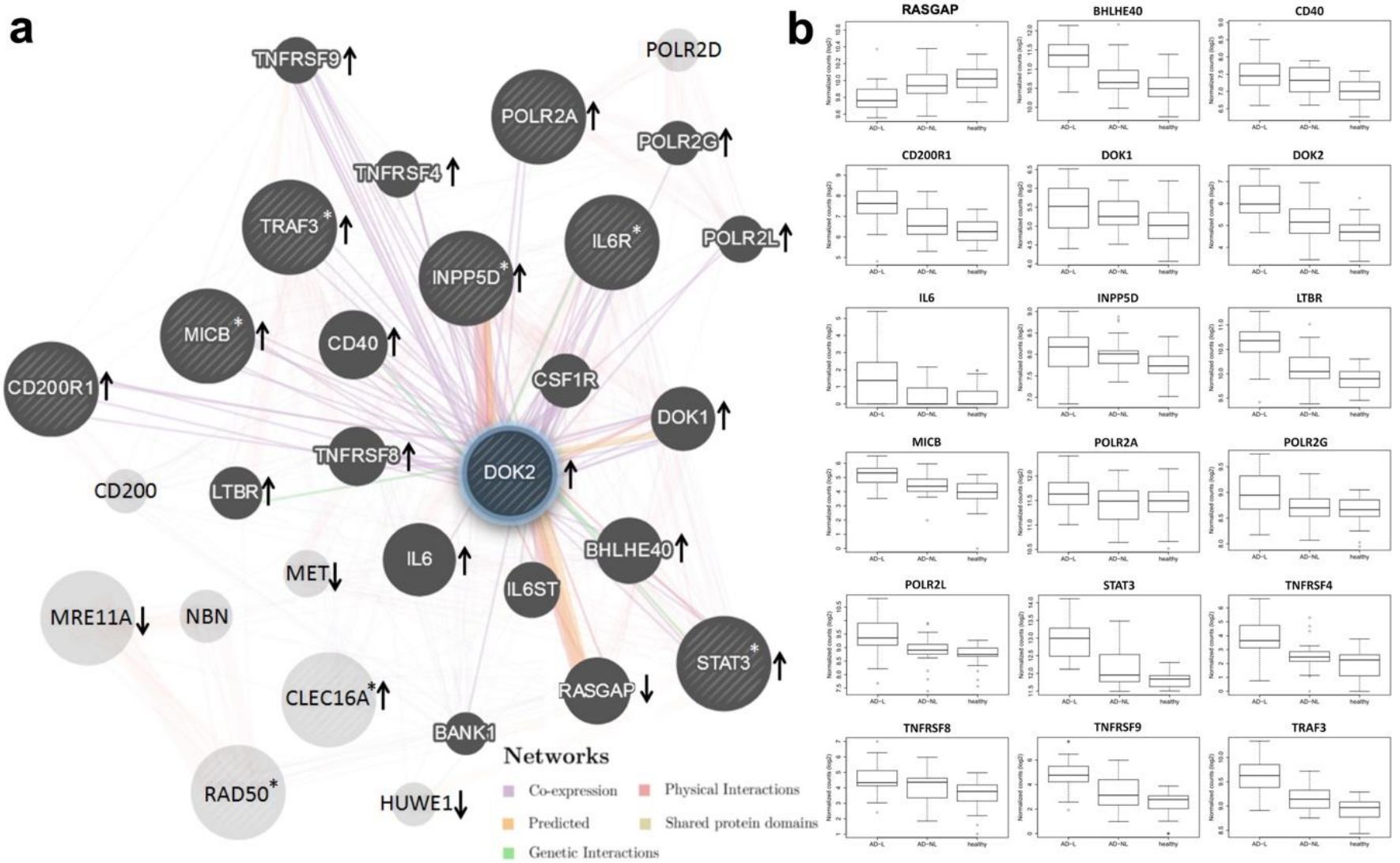
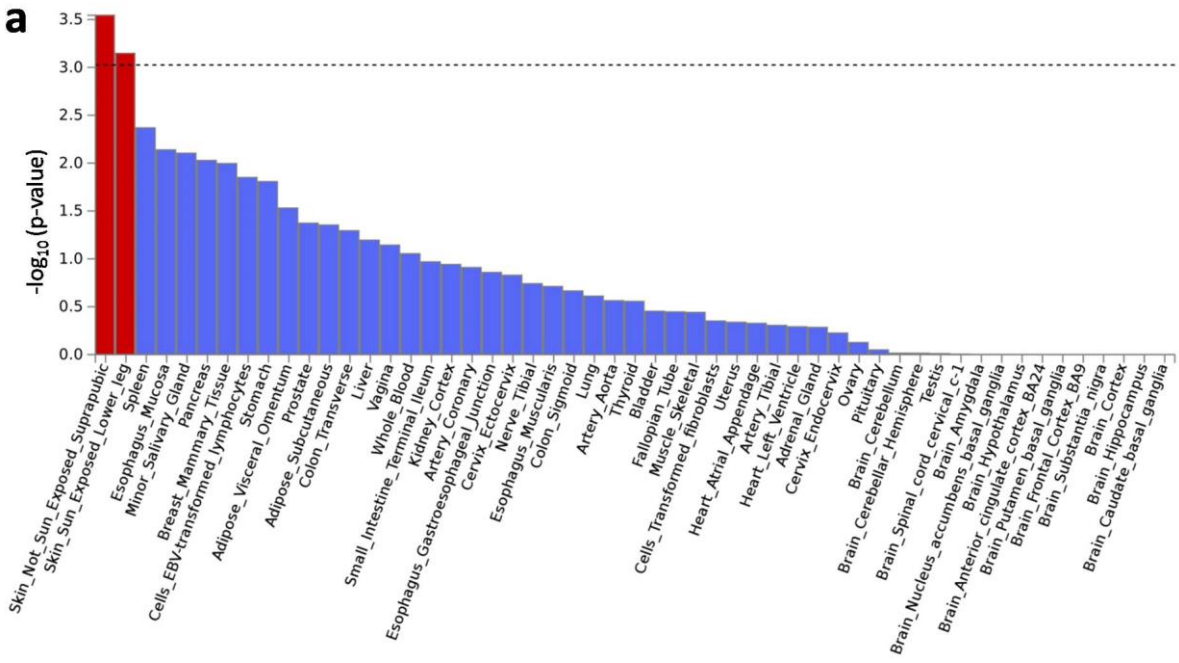
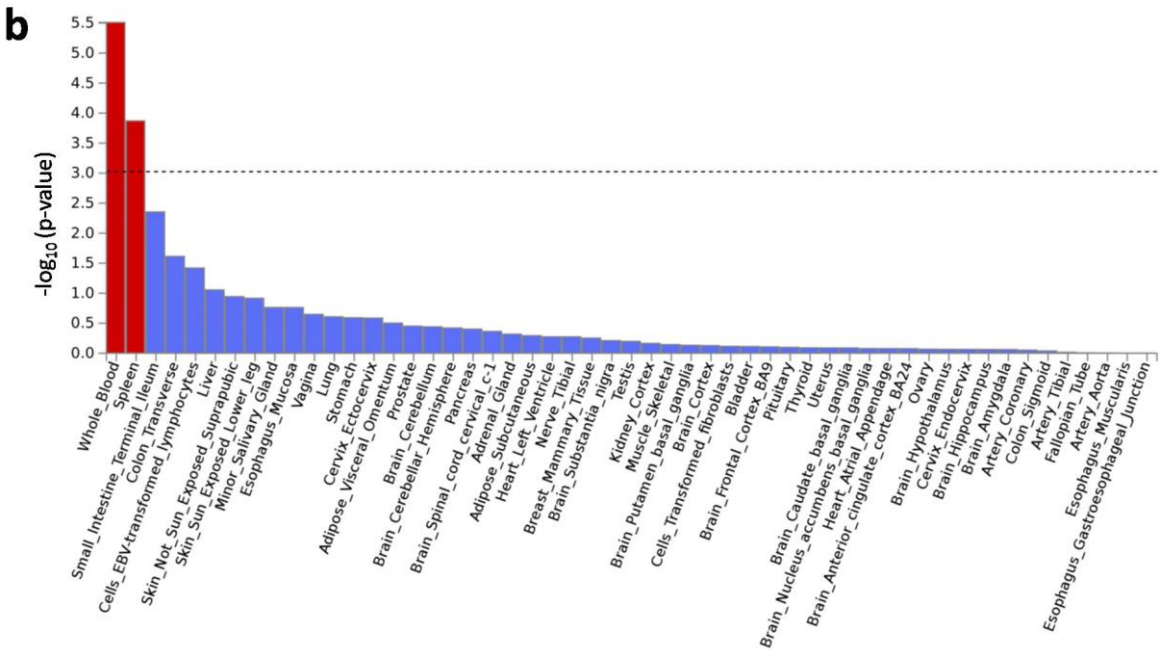

a
TNFRSF9
POLR2D
POLR2A
POLR2G
TNFRSF4
TRAF3*
IL6R*
POLR2L
INPP5D*
MICB*
CD40
CD200R1
CSF1R
DOK1
TNFRSF8
DOK2
CD200
LTBR
IL6
BHLHE40
MET
IL6ST
MRE11A
NBN
STAT3*
CLEC16A*
RASGAP
BANK1
RAD50*
HUWE1
Networks
Co-expression
Physical Interactions
Predicted
Shared protein domains
Genetic Interactions
b
RASGAP
BHLHE40
CD40
CD200R1
DOK1
DOK2
IL6
INPP5D
LTBR
MICB
POLR2A
POLR2G
POLR2L
STAT3
TNFRSF4
TNFRSF8
TNFRSF9
TRAF3
AD-L
AD-NL
healthy

Figure 4 – Tissue specificity analysis using 53 tissue types from GTEx identifies a significant positive relationship between skin tissue gene expression profiles and exome chip association statistics. Exome chip single variant association score statistics (exome chip data set) and GWAS SNP association summary statistics from the EAGLE Eczema Consortium (GWAS data set), respectively, were converted to AD-gene association Z scores using a multiple linear principal components regression model as implemented in MAGMA(30) thereby ensuring that LD between SNPs is accounted for (*see in* ***Online Repository Methods***). We conducted tissue specificity analysis with FUMA(42) to test for a positive relationship between tissue-specific gene expression profiles constructed from 53 tissue types and 11,688 samples from GTEx v7 and AD-gene associations (either from exome chip or GWAS data set) represented by gene-based Z scores (*see in* ***Online Repository Methods***). We used Bonferroni correction $p < 0.05/53 = 9.43 \times 10^{-4}$ as significance cut-off. (**a)** The exome chip data set revealed a positive relationship between gene expression profiles from skin tissues "sun exposed" and "non-sun exposed" and AD-gene associations. **(b)** In comparison, the GWAS data set showed an association with tissues "whole blood" and "spleen".




b
-log10 (p-value)
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Whole_Blood
Spleen
Small_Intestine_Terminal_Ileum
Colon_Transverse
Cells_EBV-transformed_lymphocytes
Liver
Skin_Not_Sun_Exposed_Suprapubic
Skin_Sun_Exposed_Lower_leg
Minor_Salivary_Gland
Esophagus_Mucosa
Vagina
Lung
Stomach
Cervix_Ectocervix
Adipose_Visceral_Omentum
Prostate
Brain_Cerebellum
Brain_Cerebellar_Hemisphere
Brain_Spinal_cord_cervical_c-1
Pancreas
Adrenal_Gland
Adipose_Subcutaneous
Heart_Left_Ventricle
Nerve_Tibial
Breast_Mammary_Tissue
Brain_Substantia_nigra
Testis
Kidney_Cortex
Muscle_Skeletal
Brain_Putamen_basal_ganglia
Brain_Cortex
Cells_Transformed_fibroblasts
Bladder
Brain_Frontal_Cortex_BA9
Pituitary
Thyroid
Uterus
Brain_Caudate_basal_ganglia
Brain_Nucleus_accumbens_basal_ganglia
Heart_Atrial_Appendage
Brain_Anterior_cingulate_cortex_BA24
Ovary
Brain_Hypothalamus
Cervix_Endocervix
Brain_Hippocampus
Brain_Amygdala
Artery_Coronary
Colon_Sigmoid
Artery_Tibial
Fallopian_Tube
Artery_Aorta
Esophagus_Muscularis
Esophagus_Gastroesophageal_Junction